9К

УДК 004.522

И.С. Кипяткова, А.А. Карпов
Санкт-Петербургский институт информатики и автоматизации РАН,
г. Санкт-Петербург, Россия
{kipyatkova; karpov}@iias.spb.su

Модуль фонематического транскрибирования для системы распознавания разговорной русской речи

Статья описывает модуль автоматического фонематического транскрибирования, создающий как базовые, так и альтернативные, характерные для разговорного стиля речи, транскрипции словоформ. Для генерации альтернативных транскрипций используются правила учета редукции и ассимиляции звуков речи. В статье представлены результаты экспериментов по формированию базовых фонематических транскрипций слов из исходного текста и оценка примененного метода создания альтернативных транскрипций.

Для функционирования любой системы распознавания речи необходим фонетический или фонематический словарь слов, используемых в системе. Для автоматического его создания авторами был разработан модуль фонематического транскрибирования. Базовые транскрипции создаются по фонетическим правилам русского языка [1]. Однако в разговорной речи произношение может варьироваться, и транскрипции произнесенных слов часто отличаются от базовых. Для учета вариативности произношения необходимо принимать во внимание явления редукции и ассимиляции звуков. Опираясь на результаты исследований, полученных в области экспериментальной фонетики [2], был сформулирован ряд правил, которые достаточно точно описывают возможные отклонения в фонематических транскрипциях, связанные с ассимиляцией и редукцией звуков внутри слова и на стыке двух слов [3].

На вход модуля транскрибирования поступают: набор предложений, составляющих тексты; словарь словоформ русского языка с отметкой ударного слога (слогов), используемый фонемный алфавит и фонетические правила, включающие в себя как базовые правила транскрибирования, так и правила учета явлений редукции и ассимиляции звуков. Результатом работы модуля является транскрибированный словарь словоформ, в который входят транскрипции, созданные по базовым правилам транскрибирования, и альтернативные транскрипции, отражающие произношение слов в разговорной русской речи.

1. Фонемный алфавит системы распознавания русской речи

Речевой поток представляет собой непрерывную линейную последовательность звуков. Устная, звучащая речь, отличается от письменной речи, которая фиксирует звучащую речь в буквенных обозначениях. Для передачи звучащей речи на письме используется особая запись, отличная от орфографической, – фонематическая транскрипция, т.е. запись звуков речи с учетом их позиционных изменений в речевом потоке с обозначением ударений в словах. В интересах орфоэпии в данном случае не требуется вводить разграничений между понятиями звука речи и фонемы [4].

Особенности тех или иных звуков фиксируются посредством специальных символов транскрипции. Для нужд общей фонетики требуются транскрипционные системы универсального назначения, пригодные для фиксации особенностей звучащей речи на любом языке и более или менее общепринятые в международных лингвистических кругах. Наиболее распространены системы, использующие в качестве символов фонем знаки латинской графики с добавлением, в случае необходимости, знаков из других графических систем. Таковы, например, система МФА (Международной фонетической ассоциации, International Phonetic Association) в разных её версиях и восходящая к ней система Л.В. Щербы. В этих системах имеются символы для обозначения согласных и гласных, в основном, для обозначения их дополнительных артикуляционных свойств (палатализованность, веляризованность, придыхательность, огублённость, назализованность, отодвинутость назад, продвинутость вперёд), для обозначения степеней силового ударения, тона и характера музыкального ударения и т.д.

В последние годы также была разработана система транскрипции SAMPA (Speech Assessment Methods Phonetic Alphabet) для большого числа европейских языков. В транскрипции SAMPA принято использовать только те символы, которые имеются на клавиатуре персонального компьютера. С их помощью оказывается возможным передавать фонетическую информацию по межкомпьютерной связи.

В данном исследовании в качестве фонемного алфавита используется модифицированный и адаптированный к кириллице вариант международного фонетического алфавита SAMPA [5]. В нашем варианте используются 48 фонем: 12 – для гласных звуков (с учетом ударных вариантов) и 36 – для согласных (с учетом твердости и мягкости звуков). Изменения коснулись гласных звуков, дополнительно к обычным вариантам гласных звуков добавлены варианты с ударением. Так как ударные и безударные гласные имеют значительные отличия в спектральных и временных характеристиках, то такое разделение позволяет увеличить точность моделирования речи. В табл. 1 представлен перечень базовых фонем, которые используются при распознавании русской речи. Знак [′] используется для обозначения ударения в слове, знак [`] – для обозначенияв акцентированного гласного звука (т.е. второго ударения в слове) и знак ['] – для обозначения мягкости согласных.

Одним из основных видов информации при транскрибировании является информация о положении ударения в слове, так как от этой информации, в значительной степени, зависит результат транскрибирования слова. В данном проекте была использована база данных словоформ русского языка с информацией об ударениях в словоформах, насчитывающая свыше 1 млн 700 тыс. словоформ. Такая база данных доступна через Интернет [6] и была создана в ходе проекта STARLING (руководитель проекта С.А. Старостин). При формировании этой базы данных был использован грамматический словарь А.А. Зализняка (свыше 160 тыс. базовых вариантов слов), правила словообразования слов русского языка и правила расстановки ударений в словоформах.

Таблица 1 – Перечень фонем для распознавания русской речи

Фонема	Слово	Транскрипция	Фонема	Слово	Транскрипция
/а/	*пара*	/п а′ р а/	/ц/	*цепь*	/ц э′ п'/
/а′/	*пара*	/п а′ р а/	/ч/	*чай*	/ч а′ й/
/и/	*мели*	/м' е′ л' и/	/ф/	*фарс*	/ф а′ р с/
/и′/	*мир*	/м' и′ р/	/ф'/	*физика*	/ф' и′ з' и к а/
/е/	*дерево*	/д' е′ р' е в а/	/в/	*ваза*	/в а′ з а/

Продолж. табл. 1

Фонема	Слово	Транскрипция	Фонема	Слово	Транскрипция
/е′/	*дерево*	/д' е′ р' е в а/	/в'/	*виза*	/в' и′ з а/
/ы/	*дыры*	/д ы′ р ы/	/с/	*сын*	/с ы′ н/
/ы′/	*дыры*	/д ы′ р ы/	/с'/	*сено*	/с' е′ н а/
/у/	*тулуп*	/т у л у′ п/	/з/	*запах*	/з а′ п а х/
/у′/	*тулуп*	/т у л у′ п/	/з'/	*корзина*	/к а р з' и′ н а/
/о′/	*город*	/г о′ р а т/	/ш/	*шар*	/ш а′ р/
/э′/	*цепь*	/ц э′ п'/	/щ/	*щука*	/щ у′ к а/
/п/	*пыль*	/п ы′ л'/	/ж/	*жир*	/ж ы′ р/
/п'/	*пить*	/п' и′ т'/	/х/	*хлеб*	/х л' е′ п/
/б/	*быть*	/б ы′ т'/	/х'/	*хитрый*	/х' и′ т р ы й/
/б'/	*бить*	/б' и′ т'/	/м/	*май*	/м а′ й/
/т/	*тост*	/т о′ с т/	/м'/	*мята*	/м' а′ т а/
/т'/	*тень*	/т' е′ н'/	/н/	*найти*	/н а й т' и′/
/д/	*дым*	/д ы′ м/	/н'/	*нить*	/н' и′ т'/
/д'/	*день*	/д' е′ н'/	/л/	*луч*	/л у′ ч/
/к/	*кот*	/к о′ т/	/л'/	*любовь*	/л' у б о′ ф'/
/к'/	*кит*	/к' и′ т/	/р/	*краб*	/к р а′ п/
/г/	*гусь*	/г у′ с'/	/р'/	*резать*	/р' е′ з а т'/
/г'/	*гибкий*	/г' и′ п к' и й/	/й/	*июль*	/и й у′ л'/

2. Базовые правила транскрибирования

Звуки речи произносятся в речевом потоке не изолированно: они тесно связаны друг с другом и изменяются в зависимости от качества соседних – предшествующих или последующих – звуков, а также в зависимости от места ударения. Эти изменения подчиняются определенным нормам, фонетические правила транскрибирования русскоязычных текстов описаны в литературе [1], [7], [8]. При транскрибировании возможны следующие позиционные изменения классов звуков: изменения гласных в положении под ударением, изменения гласных в предударных слогах, изменения гласных в заударных слогах, позиционные изменения согласных. Далее описаны позиционные изменения звуков, примененные в разработанной системе автоматического транскрибирования текста с учетом используемого фонемного алфавита.

В русском языке существует шесть позиций, в которых ударные гласные предстают в разных своих видах: (1) позиция в абсолютном начале слова не перед мягким согласным; (2) позиция между твердыми согласными и после твердого согласного не перед согласным; (3) позиция в абсолютном начале слова перед мягким согласным; (4) позиция после твердого согласного перед мягким согласным; (5) позиция после мягкого согласного не перед мягким согласным; (6) позиция между мягкими согласными. В табл. 2 представлены изменения ударных гласных. Знак t принимается для обозначения любого твердого согласного, включая шипящие и /ц/, знак t' – для обозначения любого мягкого согласного, включая /й/, и знак a – для обозначения любого ударного гласного.

Таблица 2 – Позиционные изменения гласных под ударением

Позиции ударных гласных					
1) a, at	2) tat, ta	3) at’	4) tat’	5) t’at, t’a	6) t’at’
/е/	/э/	/е/	/э/	/е/	/е/
/э/	/э/	/э/	/э/	/э/	/э/
/и/	-	/и/	-	/и/	/и/
/ы/	/ы/	-	/ы/	-	-
/а/	/а/	/а/	/а/	/а/	/а/
/о/	/о/	/о/	/о/	/о/	/о/
/у/	/у/	/у/	/у/	/у/	/у/

Позиционные изменения безударных гласных имеют место в разных позициях по отношению к ударному слогу: в предударных (иногда разделяют правила для первого предударного, а также второго и третьего предударных слогов) и заударных слогах. Позиционные изменения гласных в предударном слоге представлены в табл. 3, а изменения гласных в заударном слоге – в табл. 4.

Таблица 3 – Позиционные изменения гласных в предударном слоге

Гласный	Позиция				
	1) Начало слова	2) После задне-язычных	3) После парных твердых и /ц/	4) После парных мягких и /ч/, /щ/, /й/	5) После твердых шипящих /ш/, /ж/, /ц/
/е/	/ы/	/и/	/ы/	/и/	/ы/
/э/	/ы/	-	/ы/	-	/ы/
/и/	/и/	/и/	-	/и/	-
/ы/	-	-	/ы/	-	/ы/
/а/	/а/	/а/	/а/	/и/	/а/
/о/	/а/	/а/	/а/	/и/	/ы/
/у/	/у/	/у/	/у/	/у/	/у/

Таблица 4 – Позиционные изменения гласных в заударном слоге

Гласный	Позиция		
	1) После заднеязычных	2) После парных твердых и /ц/, /ш/, /ж/	3) После парных мягких и /ч/, /щ/
/е/	/е/	/е/	/и/
/э/	/э/	/э/	/э/
/и/	/и/	/ы/	/и/
/ы/	-	/ы/	-
/а/	/а/	/а/	/а/
/о/	/а/	/а/	/а/
/у/	/у/	/у/	/у/

Позиционные изменения согласных фонем происходят в следующих вариантах:

1. В конце слова звонкие шумные оглушаются, и на их месте выступают глухие шумные.

2. В положении перед глухими шумными согласными звонкие шумные согласные оглушаются, и на их месте выступают глухие шумные.

3. В положении перед звонкими шумными согласными, кроме /в/, /в'/, глухие шумные озвончаются, и на их месте выступают звонкие шумные. В положении перед звонкими шумными глухие /ц/, /ч/ озвончаются, и на их месте выступают звонкие /дз/ и /дж'/.

4. В положении перед мягкими зубными /т'/, /д'/ согласные /с/, /з/ смягчаются. Перед мягкими зубными /с'/, /з'/ согласные /с/, /з/, смягчаясь, объединяются с ними в одну фонему.

5. В положении перед мягкими зубными /т'/, /д'/, /с'/, /з'/ согласная /н/ произносится мягко.

6. В положении перед /ч/ согласная /т/ (орфогр. *т* и *д*), смягчаясь, объединяется с ним в фонему /ч/. В положении перед /ч/ согласная /с/ (орфогр. *с* и *з*), смягчаясь, объединяется с ним в одну фонему. Сочетание букв *тщ* произносится в беглой речи как /чш'/. Сочетание букв *сщ* произносится как /щ/. В положении перед /ч/, /щ/ согласная /н/ смягчается.

7. В положении перед /ш/, /ж/ зубные щелевые /с/, /з/ сливаются с ними в фонему /ш/ или /ж/ соответственно.

8. Две одинаковые согласные, идущие подряд, заменяются одной фонемой.

9. Происходят изменения многобуквенных последовательностей согласных: *лнц* → /нц/, *стн* → /сн/, *здн* → /зн/, *вств* → /ств/, *фств* → /ств/, *нтг* → /нг/, *ндг* → /нг/, *ндш* → /нш/, *дст* → /цт/, *тс* → /ц/, *хг* → /г/.

Перечисленные выше правила используются для создания базовых транскрипций слов. Однако в разговорной речи часть звуков может ассимилироваться или редуцироваться до полного исчезновения. Для учета этих явлений спонтанной речи необходимо создавать альтернативные транскрипции слов. В следующем разделе будут перечислены некоторые правила, позволяющие учесть возможную редукцию и ассимиляцию звуков речи.

3. Правила транскрибирования, учитывающие возможную редукцию и ассимиляцию звуков речи

Сформулированные в [3] правила для учета внутрисловной и межсловной редукции и ассимиляции можно условно разделить на три группы. В первую группу правил транскрибирования входят закономерности внутрисловной редукции для разговорного стиля:

Правило 1.1. Безударные гласные редуцируются до полного исчезновения, если они находятся:

a) между одинаковыми согласными (*филологический* /ф'илалаг'и′ческ'ий/ → /филлаг'и′ческ'ий/);
b) после одной из парных по глухости–звонкости согласных и перед соответствующей парной согласной (*создать* /сазда′т'/ → /сзда′т'/);
c) после мягкой согласной и перед любой согласной (*деревянный* /д'ер'ев'а′ный/ → /д'р'в'а′ный/);
d) после сонорной согласной и перед любой согласной (ракета /рак'е′та/ → /рк'е′та/);
e) после фрикативной согласной и перед любой согласной (*собака* /саба′ка/ → /сба′ка/).

Правило 1.2. Звонкие взрывные согласные редуцируются до полного исчезновения, если они находятся после гласной и перед безударной гласной (*падал* /па′дал/ → /па′ал/).

Правило 1.3. Сонорные согласные, стоящие перед /н/, редуцируются до полного исчезновения (*четырнадцать* /ч'еты′рнацат'/ → /ч'еты′нацат'/).

Правило 1.4. Фонемы /в/, /в'/, стоящие перед гласными, редуцируются до полного исчезновения (*совесть* /со′в'ес'т'/ → /со′ес'т'/).

Правило 1.5. Фонемы /ф/, /ф'/ редуцируются до полного исчезновения, если они находятся после глухой согласной (*сфабриковать* /сфабр'икава′т'/ → /сабр'икава′т'/).

Правило 1.6. Фонема /т'/, стоящая после /с'/ редуцируется до полного исчезновения (*в новостях* /в навас'т'а′х/ → /в навас'а′х/).

Правило 1.7. Фонема /д'/, стоящая после /з'/, редуцируется до полного исчезновения (*везде* /в'ез'д'е′/ → /в'ез'е′/).

Во вторую группу правил транскрибирования входят закономерности межсловной редукции для разговорного стиля:

Правило 2.1. Безударные гласные редуцируются до полного исчезновения, если они находятся:

a) между одинаковыми согласными (*мясо сырое* /м'а′са сыро′йе/ → /м'а′с сыро′йе/);
b) после одной из парных по глухости–звонкости согласных и перед соответствующей парной согласной (*степи большие* /с'т'е′пи бал'шы′йе/ → /с'т'е′п' бал'шы′йе/);
c) после мягкой согласной и перед согласной (*в поле стоит* /ф по′л'е стаи′т/ → /ф по′л' стаи′т/);
d) после сонорной согласной и перед согласной (*басня Крылова* /ба′с'н'а крыло′ва/ → /ба′с'н' крыло′ва/);
e) после фрикативной согласной и перед согласной (*афишу видел* /аф'и′шу в'и′д'ел/ → /аф'и′ш в'и′д'ел/).

Правило 2.2. Фонемы /т/ и /д/, стоящие после /с/ и /з/ соответственно, редуцируются до полного исчезновения (*хвост коровы* /хво′ст каро′вы/ → /хво′с каро′вы/).

Правило 2.3. Согласная /й/ в конце слова редуцируется до полного исчезновения, если ей предшествует безударная гласная, а следующее слово начинается с любой фонемы, кроме ударной гласной (*красный шар* /кра′сный ша′р/ → /кра′сны ша′р/).

В третью группу правил транскрибирования входят закономерности межсловной ассимиляции для разговорного стиля:

Правило 3.1. Глухие согласные /п/, /п'/, /т/, /т'/, /к/, /к'/, /ф/, /ф'/, /с/, /с'/, /ш/, /ш/, /ц/, /ч/ озвончаются перед фонемами /б/, /д/, /г/, /з/ или /ж/ (*с дороги* /здаро′г'и/).

Правило 3.2. Первая в слове гласная /и/ после всех твердых согласных переходит в фонему /ы/ (*лист ивы* /л'ист ы′вы/).

Правило 3.3. Сочетание фонем /с'т'/ в конце слова переходит в фонему /щ/, если следующее слово начинается с /ч/ (*есть чему* /е′с'т' чему′/ → /е′щ чему′/).

Правило 3.4. Если вначале слова стоит сочетание фонем /йи/, причем гласная безударная, оно переходит в фонему /ы/ в случае, если первое слово заканчивается на твердую согласную (*город в Якутии* /го′рат в йику'тии/ → /го′рат в ыку'тии/).

4. Процедура автоматического транскрибирования

Для осуществления автоматического транскрибирования текста был разработан модуль на языке программирования С++, позволяющий получить фонематическую транскрипцию из любого исходного текста или отдельных слов. Общий процесс транскрибирования текстов показан на рис. 1.

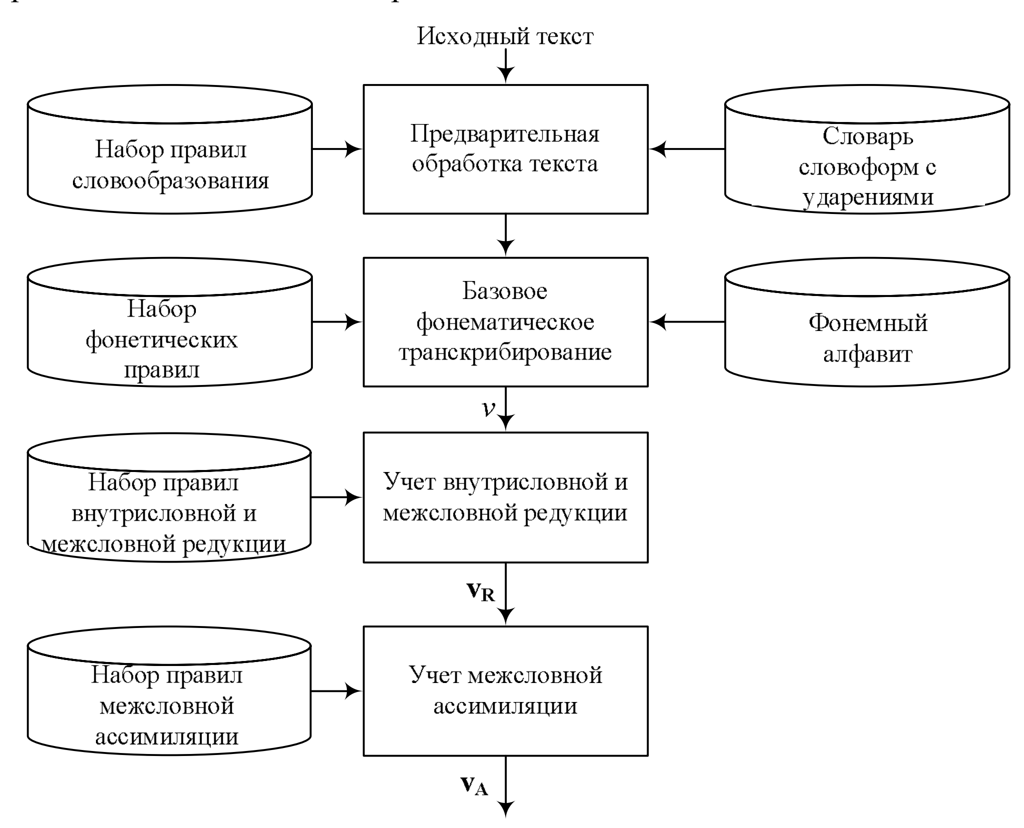


Рисунок 1 – Общий процесс автоматического транскрибирования

Перед применением фонетических правил текст предварительно обрабатывается, и в словах проставляются знаки ударения, а затем обозначаются границы основы слова путем процедуры, описанной в [9].

Вначале создаются базовые транскрипции словоформ. Процедура транскрибирования проходит последовательно за 2 одинаковых цикла, в процессе каждого из которых к анализируемому тексту последовательно применяются перечисленные выше фонетические правила в следующем порядке:

1. Согласные перед буквами *и*, *е*, *ё*, *ю*, *я* становятся мягкими.
2. Буквы *ё*, *ю*, *я* меняются на фонемы /йо/, /йу/, /йа/, если они находятся в начале слова или после гласного, в противном случае они меняются на /о/, /у/, /а/ соответственно.
3. Согласная перед *ь* становится мягкой, а сам мягкий знак убирается.
4. Парные звонкие согласные в конце слова становятся глухими.
5. Оглушаются согласные перед глухими шумными, и озвончаются согласные перед звонкими шумными.
6. Смягчаются согласные перед мягкими зубными /т'/, /д'/, /с'/, /з'/.
7. Изменения гласных под ударением (согласно табл. 2) и безударных гласных в предударном слоге (согласно табл. 3).

8. Изменения гласных в заударном слоге (согласно табл. 4).

9. Изменения двухбуквенных сочетаний согласных (включая дублирование согласных).

10. Из текста удаляется твердый знак ъ (во втором цикле).

11. В конце слов перед окончанием *о* буква *г* заменяется фонемой /в/ (например, *белого*). Для выполнения такой замены используется информация о разбиении слова на морфемы.

12. Изменения многобуквенных сочетаний согласных.

Затем полученный список базовых транскрипций поступает в блок учета внутрисловной и межсловной редукции, где для каждой базовой транскрипции слова v определяется, какие фонемы подвержены редукции. Если таких фонем в слове находится больше одной, то производится генерация всех возможных сочетаний редуцирующихся фонем. Затем каждое из полученных сочетаний обрабатывается отдельно. В результате одна альтернативная транскрипция генерируется путем удаления фонем, указанных в текущем сочетании, из базовой транскрипции. Таким образом, на выходе блока получается набор альтернативных транскрипций $\mathbf{V}_R$ данной словоформы, учитывающий все возможные сочетания редуцируемых фонем.

Далее в блоке учета межсловной ассимиляции производится анализ первых и последних фонем в транскрипции, в случае обнаружения фонем, подверженных ассимиляции, производится генерация всех возможных контекстно-зависимых вариантов транскрипций. Полученный таким образом набор транскрипций $\mathbf{V}_A$, теоретически, должен содержать все варианты произношений, которые могут возникать в разговорной речи различных людей.

5. Результаты экспериментов

Для экспериментальной проверки и создания базовых транскрипций слов использовался текст, состоящий более чем из 1000 слов, взятый из рассказов М.М. Пришвина [10]. Пример транскрибированного предложения представлен в табл. 5.

Таблица 5 – Пример транскрибирования предложения

Исходный текст	Транскрибированный текст
И это были сегодня наши последние гости	и! э!та бы!л'и СЕГОДНЯ на!шы пасл'е!дн'ийе гас'т'и!

Возникающие при транскрибировании ошибки можно разделить на 4 категории (табл. 6). Большая часть ошибок связана с неправильно проставленным ударением. Например, слово *гости* (табл. 5) было транскрибировано неверно из-за неправильного ударения (/гас'т'и!/ вместо /го!с'т'и/). Устранить данную ошибку позволит синтаксический разбор предложений с получением морфологической информации о словах, составляющих предложения. За счет этого можно избавиться от многозначности вариантов ударений в словах-омографах. В качестве синтаксического анализатора может быть использован программный модуль (вместе с исходными кодами), распространяемый по лицензии LGPL, разработанный А. Сокирко и доступный в сети Интернет [11]. Кроме того, 2 % слов не были транскрибированы, поскольку они не были найдены в словаре ударений (например, слово *сегодня* в табл. 5). Часть ошибок была связана с тем, что при удалении разделительного *ь* исчезала граница в виде фонемы /й/ между согласным и гласным. К категории «другие ошибки» относятся

следующие ошибки:

- неправильная замена буквы *г* на фонему /в/ перед окончанием *о*, поскольку данное преобразование не для всех слов является правомерным (например, *много*);
- не происходит замена буквы *г* на фонему /в/ перед *о* в середине слова (например, в слове *сегодняшний*) и т.д.

Таблица 6 – Результаты создания базовых транскрипций

Правильно транскрибированные слова	Слова, отсутствующие в словаре ударений	Слова с неправильным ударением	Исчезновение разделения при удалении разделительного *ь*	Другие ошибки
93,9 %	2,0 %	2,6 %	1,0 %	0,5 %

Результаты экспериментов показали, что базовая фонематическая транскрипция создается из исходного текста с приемлемым качеством, что позволяет использовать полученную транскрипцию для обучения системы распознавания речи и создания словаря распознавания.

Для тестирования подмодуля редукции и ассимиляции модуля автоматического транскрибирования использовался словарь, построенный по названиям рубрик электронного каталога «Желтые страницы Санкт-Петербурга». До применения правил учета внутрисловной и межсловной редукции и ассимиляции размер словаря составлял 17662 транскрипции словоформ, после применения правил учета внутрисловной и межсловной редукции количество транскрипций увеличилось на 180195, а после применения правил межсловной ассимиляции – на 12108. Размер словаря увеличился более чем на порядок, т.е. в среднем мы получаем 11 вариантов произношения каждого слова. Соотношение между базовыми транскрипциями, транскрипциями, получившимися после применения правил внутрисловной и межсловной редукции и после применения правил межсловной ассимиляции показано на рис. 2.

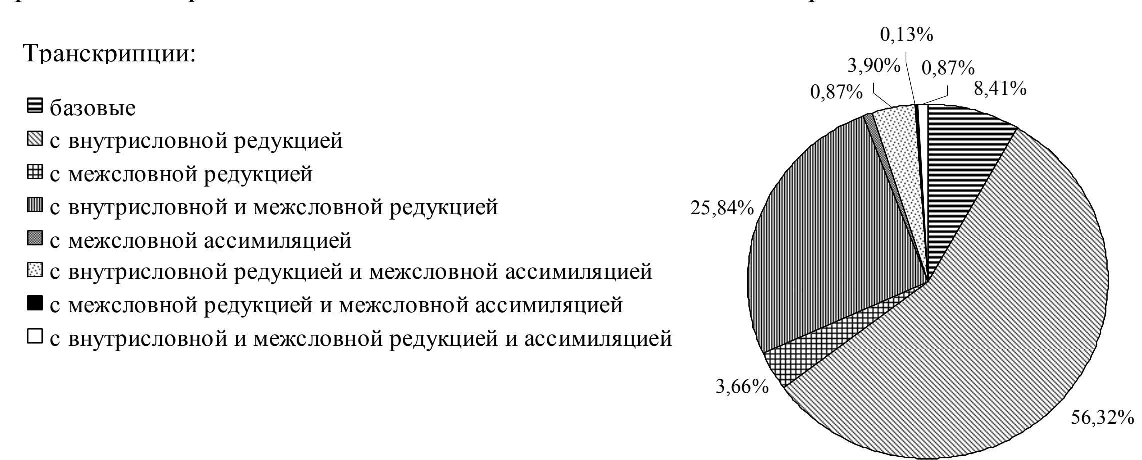


Рисунок 2 – Соотношение количества транскрипций по принципу образования

На рис. 3 представлено распределение числа транскрипций в расширенном словаре по текстовому корпусу «Желтые страницы Санкт-Петербурга». Максимальное число словоформ имеет 4 транскрипции. При этом встречаются слова с очень большим числом альтернативных транскрипций (до 2048), но у 1057 словоформ нет альтернативных транскрипций (например, *абажур*).

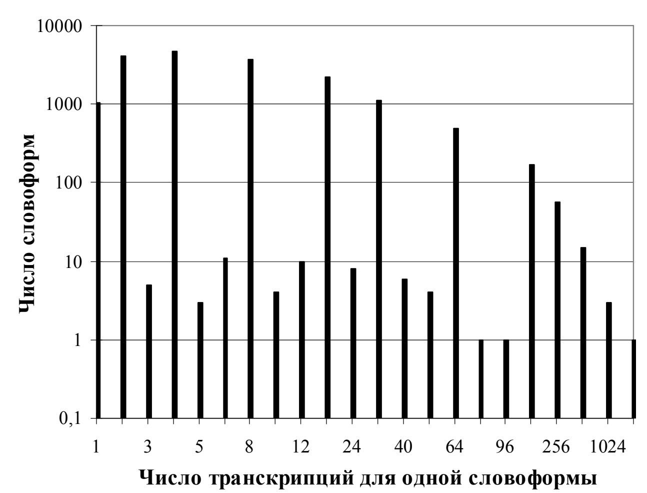


Рисунок 3 – Распределение числа транскрипций в расширенном словаре по текстовому корпусу «Желтые страницы Санкт-Петербурга»

Число альтернативных транскрипций, возникающих после применения правил редукции к одной базовой транскрипции, равно суммарному числу возможных сочетаний индексов фонем, подлежащих редукции. Таким образом, число транскрипций в расширенном словаре для любой словоформы будет кратно суммарному числу сочетаний плюс одна базовая транскрипция. Исключения составляют те словоформы, у которых к последней фонеме можно применить одновременно и правило редукции, и правило ассимиляции (словоформы, заканчивающиеся на /ф/, /ф'/ или /с'т'/), поскольку в случае редукции данной фонемы к ней не нужно применять правило ассимиляции. Однако таких словоформ немного, и именно им на гистограмме соответствуют меньшие значения (рис. 3).

Для того чтобы определить, какая часть словоформы наиболее часто подвергается изменению, для тестового словаря была определена относительная частота возникновения редукций и ассимиляций в основах и окончаниях словоформ, R_{stem} и R_{end} соответственно. Для этого использовались следующие соотношения:

$$\mathrm{R}_{\text{stem}} = \sum_{i=1}^{L} \frac{s_{stem_i}}{\mathrm{p}_{\text{stem}_i}}; \quad R_{end} = \sum_{i=1}^{K} \frac{s_{end_i}}{p_{end_i}};$$

где $\mathrm{s}_{\text{stem}_i}$ – число редукций и ассимиляций в основе *i*-ой словоформы; p_{stem_i} – число фонем в основе *i*-ой словоформы; s_{end_i} – число редукций и ассимиляций в окончании *i*-ой словоформы; p_{end_i} – число фонем в окончании *i*-ой словоформы; L – число уникальных основ (4790 для данного словаря); K – число уникальных окончаний (225 для данного словаря).

В результате произведенных расчетов относительная частота возникновения редукций и ассимиляций в основах словоформ $R_{stem} = 0{,}213$, в окончаниях – $R_{end} = 0{,}296$. Следовательно, редукции и ассимиляции чаще возникают в окончании, чем в основе.

Также был проведен анализ частоты применения каждого из правил. Результаты представлены на рис. 4. Наиболее часто использовалось правило 1.1с, связанное с редукцией безударных гласных после мягких согласных. Реже всего – правило 2.2, по которому согласные /т/ и /д/, стоящие на конце слов, редуцируются после фонем /с/ и /з/ соответственно.



Рисунок 4 – Частота применения правил редукции и ассимиляции

Указанные статистические результаты помогают оценить состав полученного расширенного словаря и провести его последующую оптимизацию с учетом специфики предметной области.

Заключение

Разработанный модуль позволяет автоматически создавать словарь базовых и альтернативных фонематических транскрипций слов для системы распознавания русской речи. Создание альтернативных транскрипций позволяет учесть такие явления разговорной речи, как редукция и ассимиляция звуков. Альтернативные транскрипции генерируются по правилам, разработанным фонетистами, однако результаты экспериментов показали, что применение всех правил приводит к значительному увеличению объема словаря и возможному созданию неправдоподобных транскрипций. Дальнейшая работа будет направлена на выставление весов для альтернативных транскрипций, что позволит отсеять редкие варианты произношения.

Работа проводится при поддержке гранта Российского Фонда Фундаментальных Исследований № 08-08-00128 «Моделирование нефонемных речевых элементов и создание альтернативных транскрипций для распознавания спонтанной русской речи», гранта Комитета по науке и высшей школе Правительства Санкт-Петербурга № 30-04/131 «Разработка бимодальной системы аудиовизуального распознавания слитной русской речи», а также Фонда Содействия Отечественной Науке.

Литература

1. Русская грамматика: В 2 т. / Редкол.: Н.Ю. Шведова (гл. ред.) и др. – М.: Наука, 1980. – 783 с.
2. Русская разговорная речь / Под ред. Е.А. Земской. – М.: Наука, 1973. – 485 с.
3. Лобанов Б.М., Цирульник Л.И. Моделирование внутрисловных и межсловных фонетико-акустических явлений полного и разговорного стилей в системе синтеза речи по тексту // Труды первого междисциплинарного семинара «Анализ разговорной русской речи» ($АР^3$ - 2007). – СПб.: ГУАП, 2007. – С. 57-71.
4. Богданова Н.В. Произношение и транскрипция: Учебно-методическое пособие по современному русскому литературному языку. – СПб.: Филологический факультет СПбГУ, 2001.
5. Режим доступа: http://www.phon.ucl.ac.uk/home/sampa/
6. Режим доступа: http://starling.rinet.ru
7. Скрелин П.А. Сегментация и транскрипция. – СПб.: Изд-во СПбГУ, 1999.
8. Тихонов А.Н., Тихонова Е.Н., Тихонов С.А. Словарь-справочник по русскому языку: правописание, произношение, ударение, словообразование, морфемика, грамматика, частота употребления слов: Ок. 26000 слов / Под ред. А.Н. Тихонова. – М.: ТОО «Словари», 1995. – 704 с.
9. Ronzhin A., Leontieva An., Kagirov I., Karpov A. Morpho-Phonetic Tree Decoder for Russian // Proc. of 12-th International Conf. on Speech and Computer SPECOM // Moscow (Russia). – 2007. – P. 491-498.

I.S. Kipyatkova, A.A. Karpov

The Phonemic Transcription Module for a System of Conversational Russian Speech Recognition

The paper describes a module of automatic phonemic transcription which creates both basic and alternative transcriptions that are typical for conversational speech. The rules of sound reduction and assimilation are used for generation of alternative transcriptions. The experimental results of formation of basic phonemic transcriptions of words from an original text and evaluation of the applied method for creation of alternative transcriptions are presented in the paper.

Статья поступила в редакцию 18.07.2008.